09/763149

Claims

1.   A method for the co-articulation-specific concatenation
of audio segments, in order to generate synthesised acoustical
data which reproduces a sequence of concatenated sounds/
phones, comprising the following steps:
- selecting at least two audio segments which contain bands,
each of which reproducing a portion of a sound/phone or a
portion of a sound/phone sequence,
- establishing a band to be used of an earlier audio segment;
- establishing a band to be used of a later audio segment,
which begins with the later audio segment and ends with the
co-articulation band of the later audio segment which follows
the initially used solo articulation band;
- with the duration and position of the bands to be used being
determined as a function of the earlier and later audio seg-
ments; and
- concatenating the established band of the earlier audio seg-
ment with the established band of the later audio segment, in
that the instance of concatenation, as a function of proper-
ties of the used band of the later audio segment, is set in a
band which begins immediately before the used band of the
later audio segment and ends with same.

2.   The method according to Claim 1, characterised in that
- the instance of concatenation is set in a band which lies in
the vicinity of the boundaries of the initially to be used
solo articulation band of the later audio segment, if the band
of same to be used reproduces a static sound/phone at the be-
ginning; and
- a downstream portion of the band to be used of the earlier
audio segment and an upstream portion of the band to be used
of the later audio segment are processed by means of suitable
transfer functions and added in an overlapping manner (cross

fade), with the transfer functions and the length of an over-lapping portion of the two bands being determined depending on the audio segments to be concatenated.

5   3.   The method according to Claim 1 or 2, characterised in that
- the instance of concatenation is set in a band which lies immediately before the band to be used of the later audio segment, if the used band of same reproduces a dynamic sound/
10   phone at the beginning; and
- a downstream portion of the band to be used of the earlier audio segment and an upstream portion of the band to be used of the later audio segment are processed by means of suitable transfer functions and joined in a non-overlapping manner
15   (hard fade), with the transfer functions being determined depending on the acoustical data to be synthesised.

    4.   The method according to one of Claims 1 to 3, character-ised in that for a sound/phone or a portion of the sequence of
20   concatenated sounds/phones at the start of the concatenated sound/phone sequence a band of an audio segment is selected so that the start of the band reproduces the properties of the start of the concatenated sound/phone sequence.

25   5.   The method according to one of Claims 1 to 4, character-ised in that for a sound/phone or a portion of the sequence of concatenated sounds/phones at the end of the concatenated sound/phone sequence a band of an audio segment is selected so that the end of the band reproduces the properties of the end
30   of the concatenated sound/phone sequence.

    6.   The method according to one of Claims 1 to 5, character-ised in that the voice data to the synthesised is combined in groups, each of which being described by an individual audio
35   segment.

7.   The method according to one of Claims 1 to 6, character-
ised in that an audio segment is selected for the later audio
segment band, which reproduces the highest number of success-
5   ive portions of the sounds/phones of the sound/phone sequence,
in order to use the smallest number of audio segment bands in
the generation of the synthesised acoustical data.

8.   The method according to one of Claims 1 to 7, character-
10   ised in that a processing of the used bands of individual
audio segments is carried out by means of suitable functions
depending on properties of the concatenated sound/phone
sequence, with these properties involving i.a. a modification
of the frequency, the duration, the amplitude, or the spec-
15   trum.

9.   The method according to one of Claims 1 to 8, character-
ised in that a processing of the used bands of individual
audio segments is carried out by means of suitable functions
20   in a band, in which the instance of concatenation lies, with
these functions involving i.a. a modification of the frequen-
cy, the duration, the amplitude, or the spectrum.

10.   The method according to one of Claims 1 to 9, character-
25   ised in that the instance of concatenation is set in places of
the bands to be used of the earlier and/or later audio seg-
ment, in which the two used bands are in agreement with re-
spect to one or several suitable properties, with these pro-
perties including i.a.: zero point, amplitude values, gradi-
30   ents, derivatives of any degree, spectra, tone levels, amplit-
ude values within a frequency band, volume, style of speech,
emotion of speech, or other properties covered in the phone
classification scheme.

11.   The method according to one of Claims 1 to 10, character-
35   ised in that

- the selection of the used bands of individual audio segments, their processing, their variation, as well as their concatenation are additionally carried out with the application of heuristic knowledge which is obtained by an additionally carried out heuristic method.

12. The method according to one of Claims 1 to 11, characterised in that
- the acoustical data to be synthesised is voice data, and the sounds are phones.

13. The method according to one of Claims 2 to 12, characterised in that
- the static phones include vowels, diphtongs, liquids, vibrants, fricativen and nasals.

14. The method according to one of Claims 3 to 13, characterised in that and
- the dynamic phones include plosives, affricates, glottal stops, and click sounds.

15. The method according to one of Claims 1 to 14, characterised in that
- a conversion of the synthesised acoustical data to acoustical signals and/or voice signals is carried out.

16. A device for the co-articulation-specific concatenation of audio segments, in order to generate synthesised acoustical data which reproduces a sequence of phones, comprising:
- a database (107) in which audio segments are stored, each of which reproducing portion of a phone or portions of a sequence of (concatenated) phones;
- and/or any upstream synthesis means (108) which supplies audio segments;

- a means (105) for the selection of at least two audio segments from the database (107) and/or the upstream synthesis means (108); and
- a means (111) for the concatenation of audio segments,

5 characterised in that the concatenation means (111) is suited for
- defining a band to be used of an earlier audio segment;
- defining a portion to be used of a later audio segment in a band which starts with the later audio segment and ends after

10 a co-articulation band of the later audio segment, which follows after the initially used solo articulation band;
- determining the duration and position of the used bands depending on the earlier and later audio segments; and
- concatenating the used band of the earlier audio segment

15 with the used band of the later audio segment by defining the instance of concatenation as a function of properties of the used band of the later audio segment in a band which starts immediately before the used band of the later audio segment and ends with same.

20

17. The device according to Claim 16, characterised in that the concatenation means (111) comprises:
- means for the concatenation of the used band of the earlier audio segment with the used band of the later audio segment,

25 whose used band reproduces a static phone at the beginning in the vicinity of the boundaries of the initially occurring solo articulation band of the used band of the later audio segment;
- means for processing a downstream portion of the used band of the earlier audio segment and an upstream portion of the

30 used band of the later audio segment by suitable transfer functions; and
- means for the overlapping addition of the two bands in an overlapping portion (cross fade), which depends on the audio segments to be concatenated, with the transfer functions and

the length of an overlapping portion of the two bands being determined depending on the acoustical data to be synthesised.

18. The device according to Claim 16 or 17, characterised in that the concatenation (111) means comprises:
- means for the concatenation of the used band of the earlier audio segment with the used band of the later audio segment, whose used band reproduces a dynamic phone at the beginning, immediately before the used band of the later audio segment;
- means for processing a downstream portion of the used band of the earlier audio segment and an upstream portion of the used band of the later audio segment by suitable transfer functions, with the transfer functions being determined depending on the acoustical data to be synthesised; and
- means for the non-overlapping joining of the two audio segments.

19. The device according to one of Claims 16 to 18, characterised in that the database (107) includes audio segments or the upstream synthesis means (108) supplies audio segments which comprise bands which at the start reproduce a phone or a portion of the concatenated phone sequence at the start of the concatenated phone sequence.

20. The device according to one of Claims 16 to 19, characterised in that the database (107) includes audio segments or the upstream synthesis means (108) supplies audio segments which comprise bands, whose ends reproduce a phone or a portion of the concatenated phone sequence at the end of the concatenated phone sequence.

21. The device according to one of Claims 16 to 19, characterised in that the database (107) includes a group of audio segments or the upstream synthesis means (108) supplies audio

segments which comprise bands, whose starts each reproduce
only a static phone.

22. The device according to one of Claims 16 to 21, charac-
terised in that the concatenation means (111) comprises:
- means for the generation of further audio segments by con-
catenation of audio segments, with the starts of the bands
each reproducing a static phone, each with a band of a later
audio segment whose used band reproduces a dynamic phone at
the start, and
- a means which supplies the further audio segments to the
database (107) or the selection means (105).

23. The device according to one of Claims 16 to 22, charac-
terised in that, in the selection of the audio segment bands
from the database (107) or the upstream synthesis means (108),
the selection means (105) is suited to select the audio seg-
ments which reproduce the greatest number of successive por-
tions of concatenated phones of the concatenated phone se-
quence.

24. The device according to one of Claims 16 to 23, charac-
terised in that the concatenation means (111) comprises means
for processing the used bands of individual audio segments
with the aid of suitable functions, depending on properties of
the concatenated phone sequence, with the functions involving
among others a modification of the frequency, the duration,
the amplitude, or the spectrum.

25. The device according to one of Claims 16 to 24, charac-
terised in that
- the concatenation means (111) comprises means for processing
the used bands of individual audio segments with the aid of
suitable functions in a band including the instance of conca-

tenation, with this function involving i.a. a modification of the frequency, the duration, the amplitude, or the spectrum.

26. The device according to one of Claims 16 to 25, charac-
terised in that
- the concatenation means (111) comprises means for the selec-
tion of the instance of concatenation in a place in the used
bands of the earlier and/or the later audio segment, in which
the two used bands are in agreement with respect to one or
several suitable properties, with these properties including
i.a.: zero points, amplitude values, gradients, derivatives of
any degree, spectra, tone levels, amplitude values in a fre-
quency band, volume, style of speech, emotion of speech, or
other properties covered in the phone classification scheme.

27. The device according to one of Claims 16 to 26, charac-
terised in that
- the selection means (105) comprises means for the implement-
ation of heuristic knowledge which relates to the selection of
the used bands of the individual audio segments, their pro-
cessing, their variation, as well as their concatenation.

28. The device according to one of Claims 16 to 27, charac-
terised in that
- the database (107) includes audio segments or the upstream
synthesis means (108) supplies audio segments which include
bands, each of which reproducing at least a portion of a sound
or phone, respectively, a sound or phone, respectively, por-
tions of phone sequences or polyphones, respectively, or sound
sequences or polyphones, respectively.

29. The device according to one of Claims 17 to 28, charac-
terised in that
the data base (107) includes audio segments or the upstream
synthesis means (108) supplies audio segments, with a static

sound corresponding to a static phone and comprising vowels, diphtongs, liquids, vibrants, fricatives, and nasals.

30. The device according to one of Claims 18 to 29, charac-
terised in that
- the database (107) includes audio segments or the upstream synthesis means (108) supplies audio segments, with a dynamic sound corresponding to a dynamic phone and comprising plos-ives, affricates, glottal stops, and klick speech.

31. The device according to one of Claims 16 to 30, charac-
terised in that
- the concatenation means (111) is suitable to generate syn-thesised voice data by means of the concatenation of audio segments.

32. The device according to one of Claims 16 to 31, charac-
terised in that
- means (117) are provided for the conversion of the synthe-sised acoustical data to acoustical signals and/or voice signals.

33. A data carrier which includes a computer program for the co-articulation-specific concatenation of audio segments in order to generate synthesised acoustical data which reproduces a sequence of concatenated phones, comprising the following steps:
- selection of at least two audio segments which contain bands, each of which reproducing a portion of a sound/phone or a portion of a sound/phone sequence,
characterised by the steps of:
- establishing a band to be used of an earlier audio segment;
- establishing a band to be used of a later audio segment, which begins with the later audio segment and ends with the

co-articulation band of the later audio segment which follows the initially used solo articulation band;

- with the duration and position of the bands to be used being determined as a function of the earlier and later audio seg-

5   ments; and

- concatenating the established band of the earlier audio segment with the established band of the later audio segment, in that the instance of concatenation, as a function of properties of the used band of the later audio segment, is set in

10   its established band which starts immediately before the band to be used of the later audio segment and ends with same.

34. The data carrier according to Claim 33, characterised in that the computer program selects the instance of the conca-

15   tenation of the used band of the second audio segment with the used band of the first audio segment in such a manner that

- the instance of concatenation is set in a band which lies in the vicinity of the boundaries of the initially used solo articulation band of the later audio segment, if its used band

20   reproduces a static phone at the start;

- a downstream portion of the used band of the earlier audio segment and an upstream portion of the used band of the later audio segment are processed by suitable transfer functions and added in an overlapping manner (cross fade), with the transfer

25   functions and the length of an overlapping portion of the two bands being determined depending on the audio segments to be concatenated.

35. The data carrier according to Claim 33 or 34, charac-

30   terised in that the computer program selects the instance of the concatenation of the used band of the second audio segment with the used band of the first audio segment in such a manner that

- the instance of concatenation is set in a band which lies
immediately before the used band of the later audio segment,
if its used band reproduces a dynamic phone at the start;
- a downstream portion of the used band of the earlier audio

5   segment and an upstream portion of the used band of the later
audio segment are processed by suitable transfer functions and
added in a non-overlapping manner (hard fade), with the trans-
fer functions being determined depending on the audio segments
to be concatenated.

10

36.   The data carrier according to one of Claims 33 to 35,
characterised in that the computer program selects a band of
an audio segment for a phone or a portion of the sequence of
concatenated phones at the start of the concatenated phone

15 sequence, the start of which reproduces the properties of the
start of the concatenated sequence of phones.

37.   The data carrier according to one of Claims 33 to 36,
characterised in that the computer program selects a band of

20 an audio segment for a phone or a portion of the sequence of
concatenated phones at the end of the concatenated phone
sequence, the end of which reproduces the properties of the
end of the concatenated sequence of phones.

25 38.   The data carrier according to one of Claims 33 to 37,
characterised in that the computer program carries out a pro-
cessing of the used bands of individual audio segments with
the aid of suitable functions depending on properties of the
phone sequence, with the functions involving i.a. modification

30 of the frequency, the duration, the amplitude, or the
spectrum.

39.   The data carrier according to one of Claims 33 to 38,
characterised in that the computer program selects an audio

35 segment band for the later audio segment band which reproduces

the highest number of successive portions of the concatenated phones in the phone sequence, in order to use the smallest number of audio segment bands in the generation of the synthesised acoustical data.

40. The data carrier according to one of Claims 39 to 45, characterised in that the computer program carries out a processing of the used bands of individual audio segments with the aid of suitable functions in a band in which the instance of concatenation lies, with these functions involving i.a. a modification of the frequency, the duration, the amplitude, or the spectrum.

41. The data carrier according to one of Claims 33 to 40, characterised in that the computer program establishes the instance of concatenation in a place of the used bands of the first and/or the second audio segment, in which the two used bands are in agreement with respect to one or several suitable properties, with these properties including i.a.: zero points, amplitude values, gradients, derivatives of any degree, spectra, tone levels, amplitude values in a frequency band, volume, style of speech, emotion of speech, or other properties covered in the phone classification scheme.

42. The data carrier according to one of Claims 33 to 41, characterised in that the computer program carries out an implementation of heuristic knowledge which relates to the selection of the used bands of the individual audio segments, their processing, their variation, as well as their concatenation.

43. The data carrier according to one of Claims 33 to 42, characterised in that the computer program is suited for the generation of synthesised voice data, with the sounds being phones.

diately before the used band of the later audio segment and
ends with same.

48. The synthesised voice signals according to Claim 47,
characterised in that the voice signals are generated in that
- the audio segments are concatenated in an instance which
lies in the vicinity of the boundaries of the later audio
segment, if the start of this band reproduces a static sound
or phone, respectively, with the static phone being a vowel, a
diphtong, a liquid, a fricative, a vibrant, or a nasal; and
- a downstream portion of the used band of the earlier audio
segment and an upstream portion of the used band of the later
audio segment are processed by means of suitable transfer
function and both bands are added in an overlapping manner
(cross fade), with the transfer functions and the length of an
overlapping portion of the two bands being determined depend-
ing on the audio segments to be concatenated.

49. The synthesised voice signals according to Claim 47 or
48, characterised in that the voice signals are generated in
that
- the audio segments are concatenated in an instance which
lies immediately before the used band of the later audio seg-
ment, if the start of this band reproduces a dynamic sound or
phone, respectively, with the dynamic phone being a plosive,
an affricate, a glottal stop, or klick speech; and
- a downstream portion of the used band of the earlier audio
segment and an upstream portion of the used band of the later
audio segment are processed by means of suitable transfer
functions and both bands are joined in a non-overlapping
manner (hard fade), with the transfer functions being determ-
ined depending on the audio segments to be concatenated.

50. The synthesised voice signals according to one of Claims
47 to 49, characterised in that

- the first sound or the first phone, respectively, or a por-
tion of the first phone sequence or of the first polyphone,
respectively, in the sequence is generated by an audio seg-
ment, whose used band at the start reproduces the properties
5   of the start of the sequence.

51. The synthesised voice signals according to one of Claims
47 to 50, characterised in that
- the last sound or the last phone, respectively, or a portion
10   of the last phone sequence or of the last polyphone, respect-
ively, in the sequence is generated by an audio segment, whose
used band at the end reproduces the properties of the end of
the sequence.

15   52. The synthesised voice signals according to one of Claims
47 to 51, characterised in that
- the voice signals are generated in that later bands of audio
segments, beginning with the reproduction of a dynamic sound
or phone, respectively, are concatenated with earlier bands of
20   audio segments, beginning with the reproduction of a static
sound or phone, respectively.

53. The synthesised voice signals according to one of Claims
47 to 52, characterised in that
25   - such audio segments are selected which reproduce the highest
number of portions of sounds or phones, respectively, of the
sequence, in order to use the smallest number of audio segment
bands in the generation of the voice signals.

30   54. The synthesised voice signals according to one of Claims
47 to 53, characterised in that
- the voice signals are generated by the concatenation of the
used bands of audio segments which are processed with the aid
of suitable functions depending on properties of the sound se-
35   quence or phone sequence, respectively, with the functions in-

volving i.a. a modification of the frequency, the duration, the amplitude, or the spectrum.

55. The synthesised voice signals according to one of Claims
47 to 54, characterised in that
- the voice signals are generated by the concatenation of the used bands of audio segments which are processed with the aid of suitable functions depending on properties of the sound sequence or phone sequence, respectively, in an area in which the instance of concatenation lies, with these properties including i.a. a modification of the frequency, the duration, the amplitude, or the spectrum.

56. The synthesised voice signals according to one of Claims
47 to 55, characterised in that the instance of concatenation lies at a place in the used bands of the earlier and/or the later audio segment, in which the two used bands are in agreement with respect to one or several suitable properties, with these properties including i.a.: zero points, amplitude values, gradients, derivatives of any degree, spectra, tone levels, amplitude values in a frequency band, volume, style of speech, emotion of speech, or other properties covered in the phone classification scheme.

57. The synthesised voice signals according to one of Claims
47 to 56, characterised in that the voice signals are suited for a conversion to acoustic signals.

58. An acoustical, optical, magnetic, or electrical data storage which contains audio segments in order generate synthesised acoustical data by means of a concatenation of used bands of the audio segments, utilising the methods according to Claim 1, or the device according to Claim 16, or the data carrier according to Claim 33.

59. The data storage according to Claim 58, characterised in that a group of the audio segments reproduces sounds or phones, respectively, or portions of sounds or phones, respectively.

60. The data storage according to Claim 58 or 59, characterised in that a group of the audio segments reproduces phone sequences or portions of phone sequences or polyphones, respectively, or portions of polyphones.

61. The data storage according to one of Claims 58 to 60, characterised in that a group of audio segments is provided whose used bands start with a static sound or phone, respectively, with the static phones comprising vowels, diphtongs, liquids, fricatives, vibrants, and nasals.

62. The data storage according to one of Claims 58 to 61, characterised in that audio segments are provided which are suitable for the conversion to acoustical signals

63. The data storage according to one of Claims 58 to 62, which additionally contains information in order to carry out a processing of the used bands of individual audio segments with the aid of suitable functions depending on properties of the acoustical data to be synthesised, with the functions involving i.a. a modification of the frequency, the duration, the amplitude, or the spectrum.

64. The data storage according to one of Claims 58 to 63, which additionally contains information relating to a processing of the used bands of individual audio segments with the aid of suitable functions in a band in which the instance of concatenation lies, with this function involving i.a. a modification of the frequency, the duration, the amplitude, or the spectrum.

65. The data storage according to one of Claims 58 to 64, which additionally provides linked audio segments, whose instance of concatenation lies at a place of the used bands of the earlier and/or later audio segment, where both used bands

5   are in agreement with respect to one or several suitable properties with these properties being i.a.: zero points, amplitude values, gradients, derivatives of any degree, spectra, tone levels, amplitude values in a frequency band, volume, style of speech, emotion of speech, or other properties cover-

10   ed in the phone classification scheme.

66. The data storage according to one of Claims 51 to 58, which additionally contains information in the form of heuristic knowledge, which relates to the selection of the

15   used bands of the individual audio segments, their processing, their variation, as well as their concatenation.

67. Sound carrier which contains data which at least partial- ly is synthesised acoustical data which were generated

20   - by means of the method according to Claim 1, or
     - by means of the device according to Claim 16, or
     - by utilising the data carrier according to Claim 58, or
     - by utilising a data storage according to Claim 58, or
     - which are the voice signals according to Claim 47.

25

68. The sound carrier according to Claim 68, characterised in that the synthesised acoustical data is synthesised voice data.